EFFICIENT CLASSIFICATION AND SECURED DEDUPLICATION OF ENCRYPTED BIG DATA USING DEEP LEARNING TECHNIQUES

SYNOPSIS

Submitted by

MANORANJITHEM V
(Reg. No. 201507232)

in partial fulfillment for the award of the degree of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION

(DEEMED TO BE UNIVERSITY)
ANAND NAGAR, KRISHNANKOIL – 626 126

TAMIL NADU, INDIA

JULY 2023

1. Introduction

With the exponential growth of data in today's digital age, the need to effectively analyze and secure large datasets has become increasingly vital [1]. In this era of big data, numerous challenges arise in areas such as healthcare, data privacy and network security [2]. This necessitates the development of advanced techniques that can address these challenges and extract meaningful insights from vast amounts of data while ensuring security measures.

This explores three significant areas of research Binary distance measures in KNN for Chronic Kidney Disease Dataset, Deduplication of Encrypted Big Data Using LKSVM Classifier and Dual Encryption with Multiple Key Set for Security Analysis and Intrusion Detection using Hierarchical Deep Learning-based Butterfly Optimization Algorithm in Big Data Platform. Each of these areas focuses on distinct but interconnected aspects of data analysis, privacy and security in the context of big data.

The first area, Binary distance measures in KNN for Chronic Kidney Disease Dataset aims to utilize the K-Nearest Neighbors (KNN) algorithm to classify and predict chronic kidney disease based on various patient attributes [3]. By employing this algorithm, medical professionals can improve diagnosis and treatment plans, leading to more effective healthcare outcomes.

The second area, Deduplication of Encrypted Big Data Using LKSVM Classifier and Dual Encryption with Multiple Key Set for Security Analysis, focuses on the protection of encrypted big data. By utilizing Support Vector Machine (SVM) classifiers [4] and dual encryption techniques with multiple key sets, this approach ensures the privacy and security of sensitive information while allowing for efficient deduplication processes. This enables organizations to eliminate duplicate data, optimize storage resources, and maintain data integrity without compromising confidentiality.

The third area, Intrusion Detection using Hierarchical Deep Learning-based Butterfly Optimization Algorithm in Big Data Platform, addresses the critical issue of network security. It combines deep learning techniques with the Butterfly Optimization Algorithm [5] to enhance the accuracy and efficiency of intrusion detection systems. By leveraging the power of hierarchical deep learning and optimization algorithms, network administrators can detect and respond to potential threats in real-time ensuring the integrity and confidentiality of sensitive data.

By combining the insights and methodologies from these three areas, researchers and practitioners can develop a comprehensive framework for advanced techniques in security and classification within big data analytics. This integrated approach not only enhances the accuracy and efficiency of data analysis and classification but also ensures the privacy, integrity, and confidentiality of sensitive information. The following sections will delve deeper into each area, providing a detailed understanding of their techniques, challenges, and potential applications in real-world scenarios.

2. Literature Survey

Srivastava et al., [6] (2022) developed a method for forecasting CKD status using diagnostic medical data available on UCI repository that includes data pre-processing, a missing value management technique, aggregation of data, and extraction and classification. In this study, a variety of physiological parameters, as well as machine learning (ML) techniques such as train different machine learning models (SVM, KNN, Random Forest, Decision Tree, ADaBoost) with the normalized dataset They developed a performance tuning nested approach that takes into account adjusting hyper - parameters as well as determining the appropriate weights to join ensembles (Ranking Weighted Ensemble)., with 98.75% accuracy, 100% Sensitivity, 96.55% Specificity and 99.03% f1score.. The findings suggested that it could be used to develop an automated system for detecting the severity of CKD.

Jongbo et al., [7] (2020) employed two ensembles approaches namely: Bagging and Random Subspace methods on three base-learners – k Nearest Neighbours, Naïve Bayes and Decision Tree for improving the classification performance of the models. Prior to classification, data preprocessing were carried out to handle missing values and data scaling to normalize the range of independent variables. The performances of the model were evaluated with accuracy, specificity, sensitivity, kappa and ROC criteria. Empirical results on Chronic kidney disease dataset obtained from UCI machine learning repository shows that the ensemble techniques provides better result than individual base learners performances with random subspace having the best performance than bagging in most of the considered situations.100% accuracy of prediction is attained using random subspace ensemble on KNN classifier.

Alaiad et al., [8] (2020) developed classification and association rule mining techniques to construct an efficient system for predicting and diagnosing CKD and its causes using weka and SPSS as platform environments. In particular, five classification algorithms, namely, naive

Bayes, decision tree, support vector machine, K-nearest neighbour, and JRip were used to achieve the research goal. In addition, Apriori algorithm was used to discover strong relationship rules between attributes. The experiments were conducted on real medical dataset collected from hospitals and patient monitoring systems. The experiments achieved high accuracy of 98.50% for K-nearest neighbour (KNN) classifier and achieved 96.00% when using classier based on association rule (JRip). They showed that applying integrative approach by combining classification algorithms and association rule mining can significantly improve the classification accuracy and be more useful for CKD prediction.

In Bhattacharyya et al., [9] (2023), hash identities and upgraded Rivest–Shamir–Adleman (RSA) have been used to secure the IoT device's data. A four-prime integer of 512 bits makes up the proposed security algorithm. A hash signature is used to provide device authentication. An effective clustering method for sensing devices based on the node level, separation from the clusters, remaining energy, and fitness has been presented for long network life. The suggested swarm-based method determines the sensor nodes' fitness. A deep neural network- (DNN-) based resource scheduling algorithm (DNN-RSM) is meant to reduce the delay and communications overhead for IoT components in the hybrid cloud system. For optimum resource allocation, all queries originating from the cluster head are categorised using DNN based on their storage, processing, and bandwidth needs. The suggested structure delivers better outcomes, particularly regarding energy use, delay, and safety level. The results of the simulation provide credence to the concept that the proposed strategy is superior to the current system. The suggested scheme includes stringent security, decreased energy usage, decreased latency, and efficient resource utilization.

Mahesh Selvi and Kavitha [10] (2022) developed a stacked discriminative de-noising convolution auto-encoder–decoder with a two-way recommendation scheme that provides secure and efficient health data to the end-users. privacy is assured to users through the modified blowfish algorithm. For structuring the big data collected from the patient, the Hadoop transform is used. The two-way system analyzes and learns more effective features from the explicit and implicit information of the patient individually, and finally, all the learned features are fused to provide an efficient recommendation. The performance of the developed system is analyzed with different statistical metrics and compared with recent approaches. From the result analysis, it is evident that the developed system performed better than the earlier approaches.

Munirathinam et al., [11] (2020) developed a new e-healthcare system for monitoring the dead disease level by using the technologies such as IoT and Cloud with the help of deep learning approach and fuzzy rules with temporal features. In the system, the medical data is retrieved from various located patients who are utilizing the e-healthcare assisting devices. First, the retrieved and encrypted data is stored in cloud by applying a secured cloud storage algorithm. Second, the stored data can be retrieved the data as original data by applying the decryption process. Third, a new cloud framework is introduced for predicting the status of heart beat rates and diabetes levels by using the medical data that is created by applying the UCI Repository dataset. In addition, a new deep learning approach which applies the Convolutional Neural Network for predicting the disease severity. The experimental results are obtained by conducting various experiments for the proposed model by using the dataset and the hospital patient records. The developed model outperformed the available disease prediction systems in terms of prediction accuracy.

Ponmalar and Dhanakoti [12] (2022) devised a new DL method for ID in big data. Primarily, the big data structure with the testbed was established for big data processing and analytics where the required data from huge dimensions were derived. The data were saved in Hadoop Distributed File Structure, which decreases quantity of latency in distributed processes. Lastly, the ID from big data was executed through CNN-related Hybrid Whale Tabu Optimization process and it efficiently distinguishes intrusion and non-intrusion.

Kan et al. [13] (2021) devised a new IoT network ID method related to Adaptive PSO-CNN (APSO-CNN). Specifically, the PSO system with the changing inertia weight was utilized to boost structured factors of 1D-CNN. The simulation results showed that the multi-type IoT network intrusion attack detection task based on APSO-CNN algorithm is effective and reliable.

Ponmalar and Dhanakoti [14] (2022) presented a novel technique to enhance the intrusion detection process by handling the fundamental big data complexities associated with different forms of heterogeneous security data. To achieve the earlier objective, the ensemble Support Vector Machine (SVM) is integrated with the Chaos Game Optimization (CGO) algorithm. The developed methodology improves the intrusion classification accuracy and also identifies nine different types of attacks present in the UNSW-NB15 dataset. The efficiency of the proposed methodology is evaluated using statistical analysis and different performance metrics such as

precision, recall, F1-score, accuracy, ROC curve, and confusion matrix by comparing it with different baseline models. The developed methodology obtained an accuracy of 96.29% when compared to the chi-SVM (89.12%) and an improvement of 6.47% is noted in the proposed methodology in terms of accuracy when compared with the chi-SVM

Ding et al., [15] (2020) developed a fuzzy and real coded chemical reaction optimization-based cluster analysis approach with feature selection for the intrusion detection system in a Big Data platform. The proposed cluster analysis model is achieved through Fuzzy C-Mean (FCM) with real-coded chemical reaction optimization, which boosts FCM to start with optimized cluster centers. Also, the use of the Flexible Mutual Information Feature Selection approach helps this model to avoid the processing of a large number of features, which drastically affects processing elements in Table 1

Table 1: Summary of Research Gap

Ref. No	Authors	Methodology/ Algorithm	Findings	Gaps
[6]	Srivastava et al. (2022)	Ensemble Learning	Improved classification accuracy in chronic kidney disease classification	Difficulty in decision making process
[7]	Jongbo et al.(2020)	Ensemble Approach	Enhanced chronic kidney disease diagnosis using ensemble techniques	Increased complexity, computation cost
[8]	Alaiad et al. (2020)	Classification and Association Rule Mining	Successful prediction of chronic kidney disease using classification and association rule mining	Computationally expensive and time consuming
[9]	Bhattacharyya et al. (2023)	IoT-Enabled Intelligent and Secure	Development of an intelligent and secure manufacturing model	Need much more security

		Blockchain in Hybrid Cloud Communication	in hybrid cloud communication	
[10]	Mahesh Selvi and Kavitha (2022)	Privacy-Aware Deep Learning Framework	Privacy-aware deep learning framework for health recommendation system on big data analysis	Need for improvement of security
[11]	Munirathinam, Sannasi Ganapathy, Kannan (2020)	Cloud and IoT based Privacy Preserved method,Secured cloud Storage Algorithm	Privacy-preserving e-Healthcare system using secured storage algorithm and deep learning	Need for new cryptographic algorithm 864 instead of ECC and also can consider the temporal 865 features in CNN for making decision effectively
[12]	Ponmalar and Dhanakoti (2022)	Hybrid Whale Tabu Algorithm,Convolutional Neural Network (CNN)	Improved intrusion detection in big data using optimized CNN architecture	Convergence problem
[13]	Kan et al.(2021)	Adaptive Particle Swarm Optimization,Convolutional Neural Network (CNN)	Novel IoT network intrusion detection approach based on adaptive PSO and CNN	In view of the heterogeneous problem of different platforms, how to merge and filter the existing network attack data sets and select the effective

				features is needed
[14]	Ponmalar and Dhanakoti (2022)	Ensemble Support Vector Machine,Chaos Game Optimization (CGO)	Effective intrusion detection using ensemble SVM and CGO in big data platform	Time consuming and computationally expensive
[15]	Ding et al. (2020)	Fuzzy and Real-coded Chemical Reaction	Intrusion detection in industrial big data environment using optimized chemical reaction	Lack of specific algorithm mentioned

3. Objectives

1. To analyze the Chronic Kidney Disease (CKD) dataset and apply Binary distance measures in k-nearest neighbors (KNN) algorithm with ten distance function for accurate classification and prediction of CKD.
2. To propose a Linear Kernel Support Vector Machine (LKSVM) Classifier and dual encryption (DE) with multiple Key Set for securing encrypted big data deduplication.
3. To develop an Intrusion Detection Approach using the Hierarchical Deep Learning-based Butterfly Optimization algorithm (ID-HDLBOA) in a Big Data Platform.

4. Research Problem

The research problem addressed in this research involves the challenges associated with data analysis, intrusion detection, and secure deduplication in the context of big data. Specifically, there is a need for accurate classification and prediction of Chronic Kidney Disease (CKD), efficient identification of intrusions in a Big Data Platform, and secure deduplication of encrypted big data. The existing methods in these areas face limitations in terms of classification accuracy, intrusion detection efficiency, and data security. Therefore, this research aims to develop advanced techniques, such as the k-nearest neighbors algorithm, the Hierarchical Deep Learning-based Butterfly Optimization algorithm, and Linear Kernel Support Vector Machine

Classifier combined with dual encryption, to overcome these challenges and provide effective solutions for improved analysis, detection, and security in big data environments.

5. Organization of the thesis

Thesis organization is given as follows:

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Binary distance measures in KNN for Chronic Kidney Disease

Chapter 4: Deduplication of Encrypted Big Data Using LKSVM Classifier and Dual Encryption with Multiple Key Set for Security Analysis

Chapter 5: An Intrusion Detection Approach using Hierarchical Deep Learning-based Butterfly Optimization Algorithm in Big Data Platform

Chapter 6: Results and Discussion

Chapter 7: Conclusion and Future work

Methodology

Binary distance measures in KNN for Chronic Kidney Disease

KNN Classification uses ten distance measures for classification. It calculates the distance between the new element and other elements classes which are known. In this paper, Chronic Kidney Disease dataset is taken from UCI database which consists of 25 variables with 400 instances. In that we have continuous, nominal and binary variables. Hence nominal variables attributes such as specific gravity, albumin and sugar are taken. We convert all the nominal variables to binary and we use knn classification. k values are chosen

In the training phase, a KNN algorithm is applied and in the test phase results are displayed. Hence the dataset partitioned into 2 phases in the ratio of 80: 20. R is a free software used by data miners which develops statistical data analysis. In R, a knn (train, test, cl) function is used for classification, where cl is the class label. The main goal of this paper is to evaluate the performance of ten distance formulae when KNN is used for binary data and also to find the best value of k. Here, we assign for k values ranging from 175 to 190 and find out the resulting error rates.

Results

Figure 1 & 2 indicates the error rates for ten different binary formulae corresponding to their k values. The error rate is minimum for different k values. For k=175,176,177 the error rates are found to be minimum.

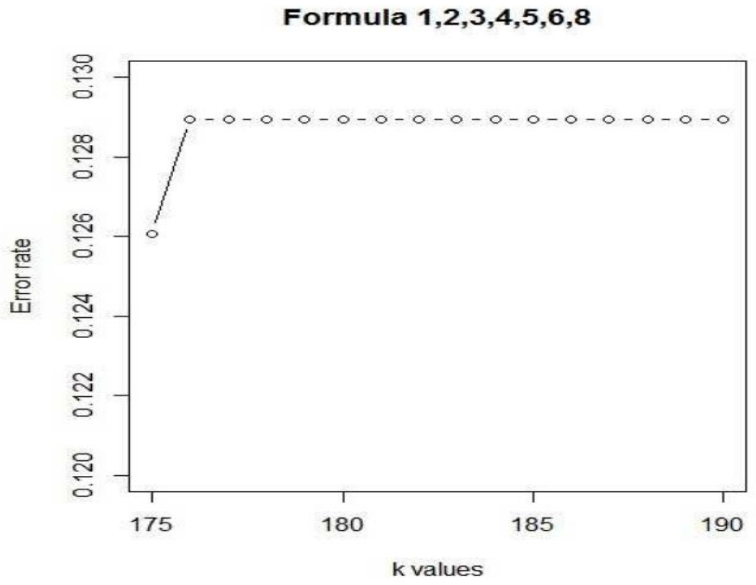

Figure 1: Formula 1,2,3,4,5,6,8

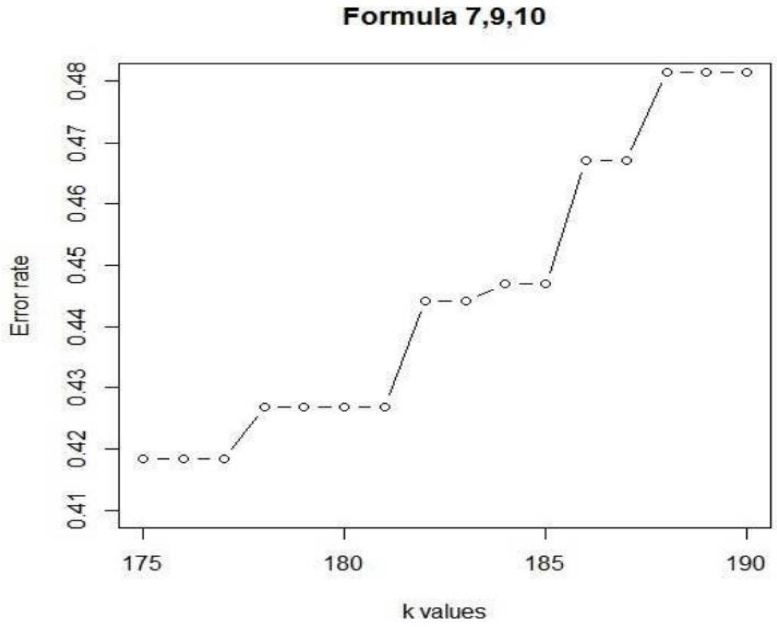

Figure 2: Formula 7,9,10

Deduplication of Encrypted Big Data Using SVM Classifier and Dual Encryption with Multiple Key Set for Security Analysis

Linear Kernel Support Vector Machine (LKSVM) Classifier and dual encryption (DE) with multiple Key Set is proposed for securing encrypted big data deduplication. Main objective of this paper is to enhance performance of deduplication of encrypted big data and to enhance the security of deduplication of encrypted big data. Firstly, the data deduplication process is executed in encrypted big data using LKSVM, and Advanced Encryption Standard (AES) and Blowfish algorithm-based DE techniques are proposed for enhancing the security of the deduplication of encrypted big data. MapReduce model is used to lessen extent of the database and discover the deduplication data from the reducer set. Deduplication is the procedure to abstain from storing similar data in multiple circumstances and the level of Deduplication achievable relies upon various elements. For this, the proposed LKSVM is developed to discover similar data to the data deduplication process. After the deduplication process is completed, to enhance security of deduplication of big data, dual encryption with Multiple Key Sets (MKS) is proposed. In dual encryption, two encryption phases are executed. In the first phase, Advanced Encryption Standard (AES) is proposed and utilized to encrypt data on the basis of private key and public key. In second phase, further to enhance the security, the second encryption technique is employed. I.e. the encrypted data is encrypted again with MKS process by using the blowfish algorithm. In the cloud, encrypted data stored to analyze duplication process based on encrypted keys, if an uploaded data or file contains a commonly-used file, the cloud server will perform the proposed data deduplication mechanism on it. Due to authorized data holders attain symmetric keys utilized for data decryption, encrypted data secures access only. Otherwise, the server will store the received data and none can learn anything about the received data.

Results

Figure 3 shows the mappers Vs LKSVM data deduplication process. The sensitivity, specificity, and accuracy can be distinguished to analyze efficiency of system. Proposed technique has achieved accuracy, sensitivity, and specificity as 79.01%, 84.41 % and, 82.0% for mappers.

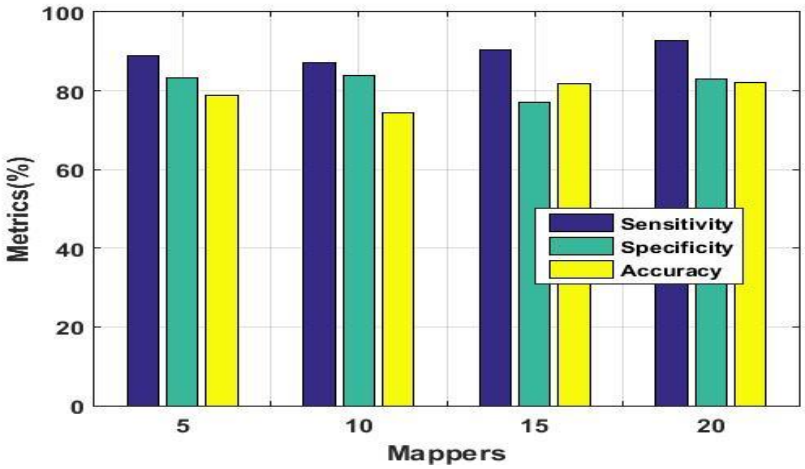

Figure 3: Classification Results for Proposed Model

Figure 4 indicates comparative analysis of classification approaches such as Neural Network (NN), Multi Kernel Support Vector Machine (MKSVM), Support Vector Machine (SVM) [31], and proposed LKSVM. The measures such as sensitivity, specificity, and accuracy can be distinguished to analyze efficiency of system. Through graphical analysis, proposed LKSVM has attained more sensitivity, specificity, and accuracy in big data classification.

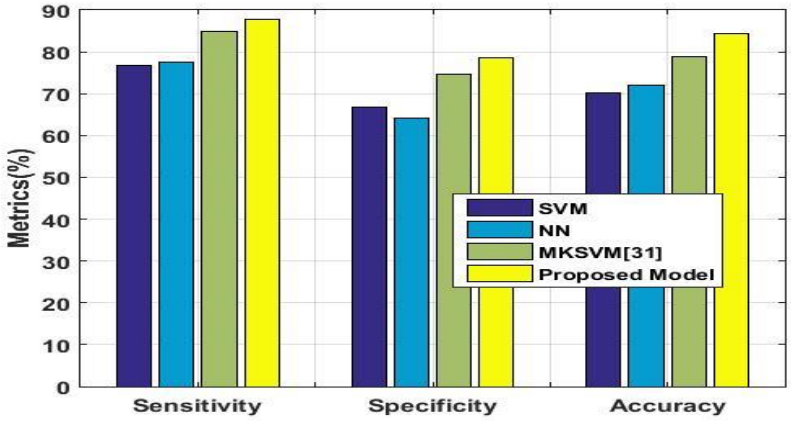

Figure 4: Comparative Analysis of Classification Approaches

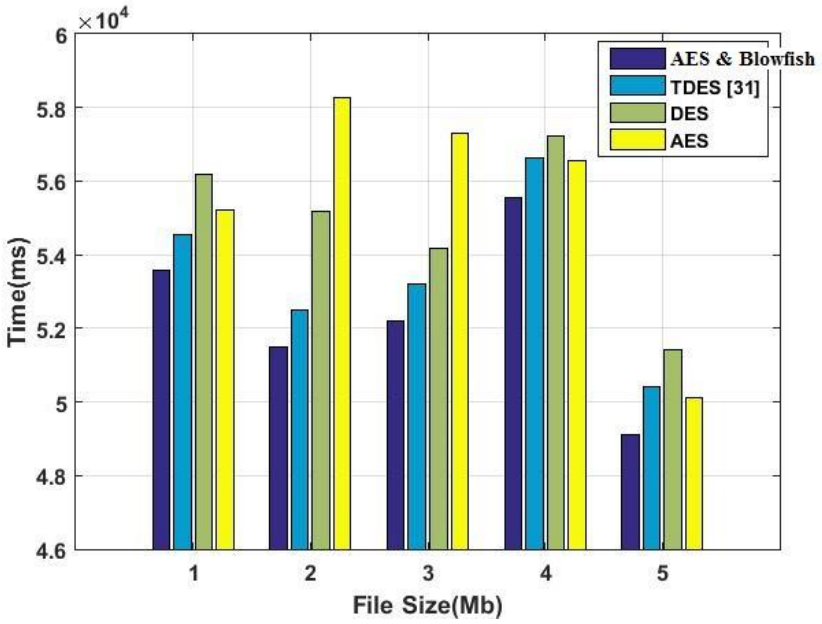

Figure 5: Comparative analysis for encryption techniques

Figure 5 indicates comparative analysis of AES and Blowfish-based proposed Data Encryption techniques and other existing techniques like TDES, DES and AES. The proposed encryption technique needs minimal time for encrypting medical data when compared to TDES , DES, and AES. But other methods like TDES, DES, and AES have taken encryption time as 54500 ms, 56100 ms and 55000 ms for file size 1 Mb. Therefore, the proposed technique has been outperformed than other methods in terms of encryption time.

An Intrusion Detection Approach using Hierarchical Deep Learning-based Butterfly Optimization Algorithm in Big Data Platform

This study develops an Intrusion Detection Approach using Hierarchical Deep Learningbased Butterfly Optimization algorithm (ID-HDLBOA) in Big Data Platform. The presented ID-HDLBOA technique combines the concept of DL with hyperparameter tuning process in Figure 6. In the presented ID-HDLBOA technique, hierarchical LSTM (Long Short Term) model is provided for intrusion detection purposes. Finally, BOA is used as a hyperparameter tuning strategy for the LSTM model and it results in improvised detection efficiency

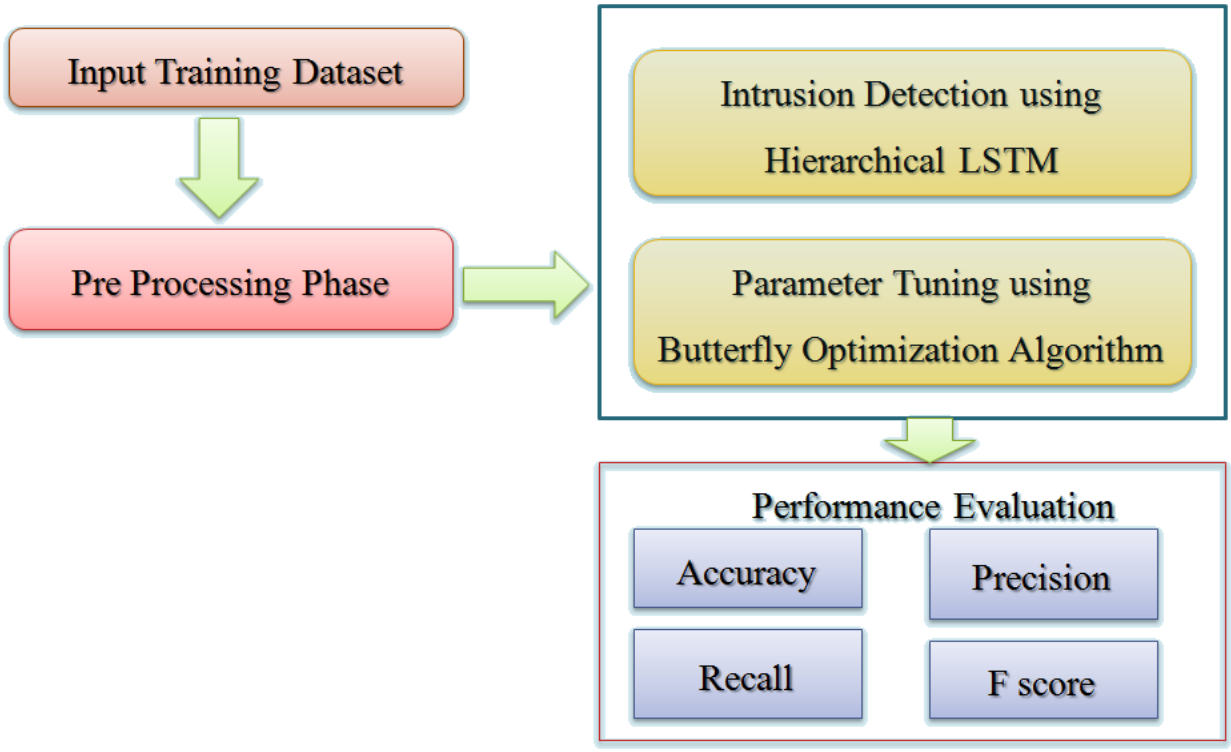

Figure 6: Flow diagram for ID-HDLBOA

Results

In Figure 7, proposed IDHDLBOA technique has gained accuracy raised of 98.55% while the DT, RF, NB, LR, and MLP models have resulted in reduced of 92.65%, 95.22%, 64.27%, 82.23%, and 77.85% respectively. Table 2 and Table 3

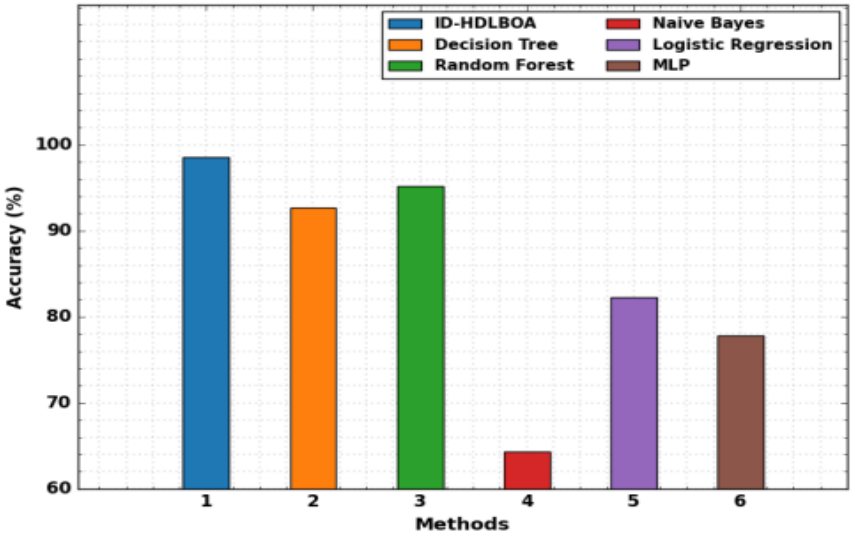

Figure 7: Accuracy analysis of ID-HDLBOA approach with other ML techniques

Table 2

INTRUSION DETECTION OUTCOME OF ID-HDLBOA APPROACH ON 80:20 OF TRS/TSS

Class	Accuracy	Precision	Recall	F-Score
Training Phase (80%)				
Normal	97.62	93.98	94.55	94.26
Attacker	98.25	96.20	95.00	95.60
Victim	99.62	99.34	98.69	99.02
Suspicious	98.00	93.94	96.27	95.09
Unknown	99.25	98.74	97.52	98.12
Average	**98.55**	**96.44**	**96.41**	**96.42**
Testing Phase (20%)				
Normal	97.00	87.18	97.14	91.89
Attacker	97.50	94.87	92.50	93.67
Victim	98.00	98.00	98.00	98.00
Suspicious	99.00	98.00	94.87	97.37
Unknown	99.50	98.00	97.44	98.70
Average	**98.60**	**96.41**	**96.39**	**96.33**

Table 3

INTRUSION DETECTION OUTCOME OF ID-HDLBOA APPROACH ON 70:30 OF TRS/TSS

Class	Accuracy	Precision	Recall	F-Score
Training Phase (70%)				
Normal	98.00	93.42	97.26	95.30
Attacker	97.57	96.27	91.49	93.82
Victim	96.57	91.04	91.04	91.04
Suspicious	98.71	97.81	95.71	96.75
Unknown	98.57	95.10	97.84	96.45
Average	**97.89**	**94.73**	**94.67**	**94.67**
Testing Phase (30%)				
Normal	96.67	84.38	98.00	91.53
Attacker	97.33	98.11	88.14	92.86
Victim	96.33	92.31	90.91	91.60
Suspicious	97.67	98.00	88.33	93.81
Unknown	98.00	92.31	98.36	95.24
Average	**97.20**	**93.42**	**93.15**	**93.01**

Conclusion and Future Enhancements

In conclusion, the three proposed works have made significant contributions to their respective fields. They have introduced innovative approaches and techniques for solving complex problems such as chronic kidney disease classification, intrusion detection in big data, and security analysis of encrypted big data. The findings and results presented have demonstrated promising outcomes in terms of accuracy, performance, and improved detection capabilities.

Future work

Future research will focus on integrating multiple algorithms or combining different techniques to develop hybrid models for chronic kidney disease classification, intrusion detection, and data security.

6. References

[1] Thomason, Jane, Big tech, big data and the new world of digital health. Global Health Journal, **5**(4) (2021) 165-168.

[2] Ahmad, Kashif, Majdi Maabreh, Mohamed Ghaly, Khalil Khan, Junaid Qadir, and Ala Al-Fuqaha. "Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges." Computer Science Review **43** (2022): 100452.

[3] Xing, Wenchao, and Yilin Bei. "Medical health big data classification based on KNN classification algorithm." IEEE Access **8** (2019): 28808-28819.

[4] Luo, Yufeng, Chongfu Zhang, Xue Wang, Xinshuai Liang, and Kun Qiu. "Robust Key Update with Controllable Accuracy Using Support Vector Machine for Secure OFDMA-PON." Journal of Lightwave Technology (2023).

[5] I.V.Pustokhina, D.A. Pustokhin, E.L. Lydia, P. Garg, A. Kadian and K. Shankar Hyperparameter search based convolution neural network with Bi-LSTM model for intrusion detection system in multimedia big data environment. Multimedia Tools and Applications, **81**(24),(2022), 34951-34968.

[6] Srivastava, Swapnita, Rajesh Kumar Yadav, Vipul Narayan, and Pawan Kumar Mall. "An Ensemble Learning Approach For Chronic Kidney Disease Classification." Journal of Pharmaceutical Negative Results (2022), 2401-2409.

[7] Jongbo, Olayinka Ayodele, Adebayo Olusola Adetunmbi, Roseline Bosede Ogunrinde, and Bukola Badeji-Ajisafe. "Development of an ensemble approach to chronic kidney disease diagnosis." Scientific African **8** (2020), e00456.

[8] Alaiad, Ahmad, Hassan Najadat, Belal Mohsen, and Khaled Balhaf. "Classification and association rule mining technique for predicting chronic kidney disease." Journal of Information & Knowledge Management **19**(01**)** (2020), 2040015.

[9] S. Bhattacharyya, S. Athithan,S. Pal,B. Sarkar,D. Akila,S. Chowdhury, K. Chandran and S. Gurusamy, An IoT-Enabled Intelligent and Secure Manufacturing Model Using Blockchain in Hybrid Cloud Communication System. Security and Communication Networks, 2023.

[10] T. Mahesh Selvi, and V. Kavitha, A privacy-aware deep learning framework for health recommendation system on analysis of big data. The Visual Computer, **38(**2)(2022), 385-403.

[11] T. Munirathinam,S. Ganapathy, and A. Kannan, Cloud and IoT based privacy preserved e-Healthcare system using secured storage algorithm and deep learning. Journal of Intelligent & Fuzzy Systems, **39**(3) (2020), 3011-3023.

[12] A.Ponmalar, and V. Dhanakoti. "Hybrid Whale Tabu algorithm optimized convolutional neural network architecture for intrusion detection in big data." Concurrency and Computation: Practice and Experience **34 (**19) (2022), e7038.

[13] Kan, Xiu, Yixuan Fan, Zhijun Fang, Le Cao, Neal N. Xiong, Dan Yang, and Xuan Li. "A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network." Information Sciences **568** (2021), 147-162.

[14] Ponmalar, A., and V. Dhanakoti. "An intrusion detection approach using ensemble support vector machine based chaos game optimization algorithm in big data platform." Applied Soft Computing **116** (2022), 108295.

[15] Ding, Weiping, Janmenjoy Nayak, Bighnaraj Naik, Danilo Pelusi, and Manohar Mishra. "Fuzzy and real-coded chemical reaction optimization for intrusion detection in industrial big data environment." IEEE Transactions on Industrial Informatics **17**, 6 (2020), 4298-4307.

7. List of publications

1. Manoranjithem.V and M. Venkatesulu "KNN Classification in Chronic Kidney Disease Dataset." International Journal of Mathematics and Computer Science 2020, 15(4), 1337–1343 .

2. Dhanasekaran, S., Anju Asokan, Arvind Kumar, C. Yamini, and Mohit Tiwari. "An Intrusion Detection Approach using Hierarchical Deep Learning-based Butterfly Optimization Algorithm in Big Data Platform." In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), 212-216. IEEE, 2023.

3. Manoranjithem.V, Dhanasekaran.S ,"Secure Deduplication of Encrypted Big Data using SVM Classifier and Dual Encryption with Multiple Key Set for Healthcare Data", Journal of Soft Computing, Springer.